

Understanding Liver Health Using the National Center for Health Statistics

Ma Somsouk · Hal F. Yee Jr. · Scott W. Biggins

Received: 2 February 2009 / Accepted: 14 April 2009 / Published online: 8 May 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract The National Center for Health Statistics (NCHS) is the principal health statistics agency for the United States. It seeks to provide accurate, relevant, and timely data on health status and utilization of health care. As such, the NCHS represents a tremendous repository of behavioral, biological, and clinical data that can be employed to identify issues and effect change in public policy related to liver health and disease. By providing an understanding of the rich, publicly available data systems within the NCHS, investigators may capitalize on an efficient means to shape current knowledge of liver disease.

Keywords Liver disease · Health statistics · CDC · NCHS · NHANES

Introduction

The first four goals of the AASLD's current strategic plan are to: (1) promote and support basic and clinical research in liver and biliary tract diseases, (2) educate health-care professionals, scientists, and the public about liver disease, (3) improve the training of professionals committed to the science and practice of hepatology, and (4) identify issues and effect change in public policy related to liver health and disease. To achieve these goals, it would be valuable to elucidate the broad impact of liver disease on health care, and personal and public health, as well as trends in liver health that are secular or in response to specific interventions or changes in public policy. In fact, collection of such vital information falls within the mission of the National Center for Health Statistics (NCHS), which is to "provide statistical information that will guide actions and policies to improve the health of the American people" (<http://www.cdc.gov/nchs/about/mission/mission.htm>). As the principal health statistics agency for the United States, the NCHS seeks to provide accurate, relevant, and timely data on health status and utilization of health care. Thus, the NCHS represents a tremendous repository of behavioral, biological, and clinical data. The aim of this report is to raise awareness of the rich, publicly available resources available through the NCHS by providing brief summaries of the most commonly used data systems. Where available, references to liver-related publications are cited at the beginning of the description of each database. We conclude by providing an example of how each database could be used to answer novel clinically important questions about hepatitis B.

M. Somsouk · H. F. Yee Jr. · S. W. Biggins
Department of Medicine, Division of Gastroenterology,
University of California, San Francisco, San Francisco,
CA, USA

M. Somsouk · H. F. Yee Jr. · S. W. Biggins
GI Health Outcomes, Policy and Economics
(HOPE) Research Program, University of California,
San Francisco, San Francisco, CA, USA

M. Somsouk · H. F. Yee Jr. · S. W. Biggins
Liver Center, University of California, San Francisco,
San Francisco, CA, USA

M. Somsouk (✉)
Division of Gastroenterology-3D, San Francisco General
Hospital, 1001 Potrero Avenue, San Francisco, CA 94110, USA
e-mail: somsoukma@medsfgh.ucsf.edu

H. F. Yee Jr.
Center for Specialty Access and Quality (CSAQ), University
of California, San Francisco, San Francisco, CA, USA

NCHS

The NCHS mainly coordinates and collects data of two types, information from population surveys and information from existing records. As seen in Table 1, there are numerous data systems available through the NCHS. A much more comprehensive table summarizing the available data systems can be accessed at: http://www.cdc.gov/nchs/data/infosheets/infosheet_summary.pdf.

The main objective of the NCHS is to present a snapshot of the health status of the US population. From a research perspective, a “snapshot” implies a cross-sectional design, which generally lacks longitudinal data; however, certain longitudinal data, such as death status, are available through updates from the National Death Index. Since longitudinal data are limited, research resulting from NCHS tends to produce association and time trend studies. By design, the individual NCHS surveys use distinct samples of the US population, that is, individual participants in one survey are not the same as those in another survey. Although each public-use dataset is deidentified, opportunities exist through the NCHS Research Data Center (RDC) for investigators to link participants to other databases that may include detailed information such as mortality data, Medicare enrollment and claims data, and Social Security benefit history (<http://www.cdc.gov/nchs/r&d/rdc.htm>).

Vital Statistics Cooperative Program

The goal of the vital statistics program is to describe birth and death patterns [1, 2]. The program obtains data from state vital registration, including all birth and death data in addition to death certificate codes. The data are stratified by year and include age, gender, and race/ethnicity, and a strength of this database stems from linking death with the primary and secondary diagnoses. A list of the data variables can be found at http://www.cdc.gov/nchs/data/dvs/file_layout.pdf. This database contains health conditions,

unlike the social security death index. Linkage with other databases is possible at the request to the RDC. Key limitations include secular trends in coding and inaccuracies in the coding of the true cause of death. Potential research applications include better understanding of life expectancy, co-morbidities, and etiologies of death, birth outcomes, and pregnancy outcomes [3].

National Health and Nutrition Examination Survey

The National Health and Nutrition Examination Survey (NHANES) describes the prevalent health conditions, health behaviors, nutrition, and environmental exposures in a sample representative of the US non-institutionalized population of the US [4–7]. Data sources are derived from an annual survey of approximately 5,000 participants of all ages with oversampling of blacks and Hispanics and those over 60 years of age. Data include household interviews, physical examinations, laboratory tests, nutritional assessment, and DNA repository.

Specific data collected that could be applicable to the study of liver disease include the following: body mass index, medications within the past 30 days, over-the-counter medications, diet, physical examination [liver ultrasound performed during the Hispanic HANES (HHANES) in 1982], laboratory data including hepatitis A, hepatitis B sAg, sAb, cAb, hepatitis C and RIBA, hepatitis D, tests of liver injury and function (excluding INR), chemistries, iron studies, HIV status, and C-reactive protein. Other laboratory data and urinalysis exist to measure environmental and toxic exposures. Several liver-related interview questions include the presence of any liver disease, age at time of disease, liver cancer, receipt of hepatitis vaccinations, alcohol use and history, illicit drug use, and sexual behavior. Note that extra DNA, serum, and plasma have been archived, and research proposals are necessary to access them. The strength of NHANES is that it provides a snapshot of the population health, and thus has been used in several studies of liver disease [4–7]. The lack of longitudinal data except for vital

Table 1 Major surveys of the National Center for Health Statistics

Surveys	Data file link
Vital Statistics Cooperative Program	http://www.cdc.gov/nchs/about/major/dvs/Vitalstatsonline.htm
National Health and Nutrition Examination Survey	http://www.cdc.gov/nchs/about/major/nhanes/datalink.htm
National Health Interview Survey	http://www.cdc.gov/nchs/about/major/nhis/quest_data_related_1997_forward.htm
National Hospital Discharge Survey	http://www.cdc.gov/nchs/about/major/hdasd/nhds.htm
National Ambulatory Medical Care Survey	http://www.cdc.gov/nchs/about/major/ahcd/ahcd1.htm
National Hospital Ambulatory Medical Care Survey	http://www.cdc.gov/nchs/about/major/ahcd/ahcd1.htm
National Home and Hospice Care Survey	http://www.cdc.gov/nchs/nnhs.htm
National Nursing Home Survey	http://www.cdc.gov/nchs/nhcs.htm

statistics and the relatively small sample size ($\sim 5,000$) are limitations to the use of this database.

Opportunities exist to petition question items for future surveys and laboratory or imaging tests as well as collaborative ancillary investigations. Upcoming laboratory data that will be collected include celiac antibody testing (IgA-tissue transglutaminase, IgA-endomysial antibody), HLA-B27, and detailed lipid profile in 2009–2010. These descriptions are not meant to be inclusive of all liver-related data available through NHANES.

National Health Interview Survey

The goal of the National Health Interview Survey (NHIS) is to describe the health status, utilization of health care, insurance, access to care, selected health conditions, immunizations, HIV testing, and health behaviors representative of the non-institutionalized population of the US [8, 9]. Like NHANES, NHIS provides a snapshot; however, it is restricted to data collected from personal interviews. NHIS is derived from sampling approximately 50,000 households. Selected supplemental interview topics are included periodically; for example, in 2007, there were questions pertaining to complementary and alternative medicine. Liver specific questions are limited, but include whether the participant has ever been told she/he ever had liver disease, liver cancer status, and immunization history [10, 11]. Liver-related questions include information on diabetes and obesity. The major strength of this database stems from the large sample size ($\sim 40,000$) and household data on prevalent conditions and health-care utilization. Limitations include the few liver-specific questions in the database; however, future supplemental topics may include more liver disease.

National Health Care Surveys

National Health Care Surveys (NHCS) are a family of provider-based surveys designed to collect information about the hospital and providers, the services rendered, and the patients they serve. They are divided into ambulatory and hospital care vs. long-term care <http://www.cdc.gov/nchs/nhcs.htm>.

1. National Hospital Discharge Survey

The goal of National Hospital Discharge Survey (NHDS) is to describe trends in hospitalizations [12, 13]. The discharge database includes 500 hospitals with over 300,000 discharges along with ICD-9 diagnosis codes and Current

Procedural Terminology (CPT) procedure codes. The strength of this database arises from details of hospitalization records that reflect the burden of disease on US hospitals. Note that the data are based on individual hospital visits, while an individual patient with repeated hospitalizations at one hospital in a calendar year is not clustered as one entry. This limitation may suggest an increased burden of disease for conditions that are associated with repeated hospitalizations. Research projects could focus on hospitalization utilization among patients with various etiologies or stages of liver disease. They may yield characteristics of associated diagnoses, conditions, or complications warranting hospitalization along with procedures performed and length of stay data [13].

2. National Ambulatory Medical Care Survey

This database is comprised of ambulatory care visits made to physician offices in the US [14–16]. These data are derived from interviews and surveys directed to 3,400 physicians spanning both primary care and subspecialists, including gastroenterologists and hepatologists, in private offices to obtain a representative sample of what providers encounter in their outpatient practice. They include ICD-9 and CPT codes and prescribing patterns, such as medications. Special focus has been placed on chronic diseases since 2005 [10]. Since ambulatory medical care in physician offices is the largest segment of health-care utilization and delivery in America, determining the utilization by disease type provides another aspect of disease impact on the US health system [17].

3. National Hospital Ambulatory Medical Care Survey

This is a physician and hospital staff survey designed to capture visits to hospital-based outpatient clinics and emergency rooms [16, 17]. It is an annual sample of 600 hospitals with 70,000 patient visits. Facility characteristics, ICD-9 and CPT codes, and prescribing patterns are included in the survey and are similar to previously described datasets above. The strength of this database is that it provides information about the population that cannot be captured by surveys of free-standing outpatient clinics.

4. National Home and Hospice Care Survey and National Nursing Home Survey

These two databases comprise data relevant to long-term health care [18, 19]. The National Home and Hospice Care Survey (NHHSS) and National Nursing Home Survey (NNHS) are derived from interviews of staff familiar with patients. The data include facility level and resident level data. Over 6,200 home health patients and another 6,200

discharged hospice patients comprise the NHHCS. Within the NNHS, 1,500 nursing homes were surveyed in 2004, which included over 18,000 residents. Long-term care provides another aspect of disease management highlighting the burden of disease towards the end of life. One key limitation of these surveys is that it has not been conducted annually.

Other Databases

This report is not meant to represent a comprehensive description of available national databases in the public domain. There are several other NCHS databases that are sampled less frequently, which are described in the summary table. Moreover, each State and County Government's public health departments collect a great deal of health status and health-care related information that is often available to the public. A link to California's Department of Public Health website is provided in Table 2. In addition, the Center for Disease Control, which is the parent institute to the NCHS, directs other databases, such as the Surveillance, Epidemiology, and End Results program (SEER), its link with Medicare, and risk factor surveillance (Table 2) [20].

These National datasets available to the public have served as the foundation for pivotal studies that have dramatically advanced the understanding of health, disease, and the delivery of health care [3, 6, 7, 10, 11, 21–23]. In an era of exponentially rising health-care costs in the US, flat or decreasing funding available for health-care researchers, such low cost-high yield data systems are an increasingly important scientific resource. To highlight a very important study that stoked the national political debate on universal health care, Wilper et al. employed the NHANES 1999–2004 to demonstrate that the uninsured population of the US relies on the more costly care provided by emergency departments for chronic disease management rather than the less expensive care provided by primary providers for the same care [23]. In the same way, progress in our understanding of the impact, prevention, and delivery of care for liver disease should be

possible using such secondary analysis of these large datasets.

Employing NCHS Databases in Hepatology Research: A Hypothetical Example

Let us say you are interested in conducting a pilot study to better understand the national burden of disease due to hepatitis B. Using the vital statistics database, you could determine the number of liver-related deaths over many years. Furthermore, you could quantify the proportion of deaths associated with hepatitis B infection. The NHANES database allows you to estimate the prevalence of hepatitis B infection across the US. Concurrently, you could assess the prevalence of surface antibody immunity against hepatitis B and document changes in immunity across age groups. Although the NHIS dataset has limited information on hepatitis B, this dataset could be used to obtain information on patients who report liver disease and liver cancers and determine their access and utilization of health-care resources. You could also use the NCHS databases to determine the changing pattern of hospitalizations and ambulatory care visits (hospital-based and free-standing clinics) related to hepatitis B infection in those who access the health-care system. Moreover, you could use this database to determine the overall utilization of health services among those infected and compare it to those without hepatitis B infection in the whole population as well as key sub-populations.

Summary

The resources available through the NCHS provide snapshots of population health and health care in the US. Opportunities for investigations specific to liver disease are great. Capitalizing on the existing data in the NCHS represents a highly efficient means to shape the current knowledge of liver disease and spur future research in novel biomarkers [24], health-related behaviors, projecting disease burden [25], health-care disparities, and much

Table 2 Other surveys beyond the National Center for Health Statistics

Surveys	Data file link
Center for Disease Control Data and Statistics Resources (Overview)	http://www.cdc.gov/DataStatistics/
Behavioral Risk Factor Surveillance System	http://www.cdc.gov/brfss/index.htm
Cancer Prevention and Control	http://www.cdc.gov/cancer/cancerburden/
Surveillance Epidemiology and End Results (SEER)	http://seer.cancer.gov/resources/ http://seer.cancer.gov/data/index.html
SEER-Medicare	http://healthservices.cancer.gov/seermedicare/
California Department of Public Health	http://www.dhs.ca.gov/

more. We hope this report will provide the reader with knowledge of large existing databases and the practical links to embark on liver-related research projects that will help to advance the science and practice of hepatology, thereby promoting liver health and optimal care of patients with liver and biliary tract diseases.

Acknowledgments We thank the individuals at the NCHS who generously shared their intimate knowledge of these data systems with us: David Lacher (NHANES), Natalie Dupree (NHANES), Geraldine McQuillan (NHANES), Susan Jack (NHIS), Nancy Sonnenfeld (NHCS), Robert Anderson (VS), and Robert Weinzimer (NCHS Outreach). We are indebted to Lyle Dennis of Cavarocchi-Ruscio-Dennis Associates who helped arrange our meeting with NCHS staff on behalf of the AASLD Public Policy Committee. This work was supported in part by the Technical Training Foundation (MS), ASGE Endoscopic Research Award (MS), the William and Mary Ann Rice Memorial Distinguished Professorship (HY), and NIH awards KL2 RR024130 from the National Center for Research Resources (SB) and DK076565 from the National Institute of Diabetes and Digestive and Kidney Diseases and from the Agency Healthcare Research and Quality (SB).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Wei R, Curtin LR, Arias E, Anderson RN. U.S. decennial life tables for 1999–2001: methodology of the United States life tables. *Natl Vital Stat Rep.* 2008;57:1–9.
- Chen CM, Yoon YH, Yi HY, Lucas DL. Alcohol and hepatitis C mortality among males and females in the United States: a life table analysis. *Alcohol Clin Exp Res.* 2007;31:285–292. doi:10.1111/j.1530-0277.2006.00304.x.
- Flegal KM, Graubard BI, Williamson DF, Gail MH. Cause-specific excess deaths associated with underweight, overweight, and obesity. *JAMA.* 2007;298:2028–2037. doi:10.1001/jama.298.17.2028.
- Ioannou GN, Weiss NS, Boyko EJ, et al. Is central obesity associated with cirrhosis-related death or hospitalization? A population-based, cohort study. *Clin Gastroenterol Hepatol.* 2005;3:67–74. doi:10.1016/S1542-3565(04)00442-2.
- Dunn W, Xu R, Wingard DL, Rogers C, Angulo P, Younossi ZM, Schwimmer JB. Suspected nonalcoholic fatty liver disease and mortality risk in a population-based Cohort study. *Am J Gastroenterol.* 2008;103:2263–2271.
- Flores YN, Yee HF, Jr., Leng M, Escarce JJ, Bastani R, Salmeron J, Morales LS. Risk factors for chronic liver disease in Blacks, Mexican Americans, and Whites in the United States: Results from NHANES IV, 1999–2004. *Am J Gastroenterol.* 2008;103:2231–2238.
- Fraser A, Longnecker MP, Lawlor DA. Prevalence of elevated alanine aminotransferase among US adolescents and associated factors: NHANES 1999–2004. *Gastroenterology.* 2007;133:1814–1820. doi:10.1053/j.gastro.2007.08.077.
- Adams PF, Lucas JW, Barnes PM. Summary health statistics for the U.S. population: National Health Interview Survey, 2006. *Vital Health Stat* 10. 2008;(236):1–104.
- Jain N, Yusuf H, Wortley PM, Euler GL, Walton S, Stokley S. Factors associated with receiving hepatitis B vaccination among high-risk adults in the United States: an analysis of the National Health Interview Survey, 2000. *Fam Med.* 2004;36:480–486.
- Ma J, Lee KV, Stafford RS. Changes in antihypertensive prescribing during US outpatient visits for uncomplicated hypertension between 1993 and 2004. *Hypertension.* 2006;48:846–852. doi:10.1161/01.HYP.0000240931.90917.0c.
- Polednak AP. Secular trend in U.S. black–white disparities in selected alcohol-related cancer incidence rates. *Alcohol Alcohol.* 2007;42:125–130. doi:10.1093/alcale/agl121.
- DeFrances CJ, Lucas CA, Buie VC, Golosinskiy A. National Hospital Discharge Survey. *Natl Health Stat Rep.* 2006;2008:1–20.
- Foreman MG, Mannino DM, Moss M. Cirrhosis as a risk factor for sepsis and death: analysis of the National Hospital Discharge Survey. *Chest.* 2003;124:1016–1020. doi:10.1378/chest.124.3.1016.
- Cherry DK, Woodwell DA, Rechtsteiner EA. National Ambulatory Medical Care Survey: 2005 summary. *Adv Data.* 2007;(387):1–39.
- Grant WC, Jhaveri RR, McHutchison JG, Schulman KA, Kauf TL. Trends in health care resource use for hepatitis C virus infection in the United States. *Hepatology.* 2005;42:1406–1413. doi:10.1002/hep.20941.
- Tsui JI, Maselli J, Gonzales R. Sociodemographic trends in national ambulatory care visits for Hepatitis C virus infection. *Dig Dis Sci.* 2008. [Epub ahead of print]
- Burt CW, McCaig LF, Rechtsteiner EA. Ambulatory medical care utilization estimates for 2005. *Adv Data.* 2007;(388):1–15.
- Gabrel CS. An overview of nursing home facilities: data from the 1997 National Nursing Home Survey. *Adv Data.* 2000;(311):1–12.
- Munson ML. Characteristics of elderly home health care users: data from the 1996 National Home and Hospice Care Survey. *Adv Data.* 1999;(309):1–11.
- Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. *Med Care.* 2002;40:IV–3–IV–18.
- Alter MJ, Kruszon-Moran D, Nainan OV, et al. The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. *N Engl J Med.* 1999;341:556–562. doi:10.1056/NEJM199908193410802.
- Coresh J, Selvin E, Stevens LA, et al. Prevalence of chronic kidney disease in the United States. *JAMA.* 2007;298:2038–2047. doi:10.1001/jama.298.17.2038.
- Wilper AP, Woolhandler S, Lasser KE, et al. Waits to see an emergency department physician: US trends and predictors, 1997–2004. *Health Aff (Millwood).* 2008;27:w84–w95. doi:10.1377/hlthaff.27.2.w84.
- Lang IA, Galloway TS, Scarlett A, et al. Association of urinary bisphenol A concentration with medical disorders and laboratory abnormalities in adults. *JAMA.* 2008;300:1303–1310. doi:10.1001/jama.300.11.1303.
- Bell BP, Kruszon-Moran D, Shapiro CN, Lambert SB, McQuillan GM, Margolis HS. Hepatitis A virus infection in the United States: serologic results from the Third National Health and Nutrition Examination Survey. *Vaccine.* 2005;23:5798–5806. doi:10.1016/j.vaccine.2005.03.060.